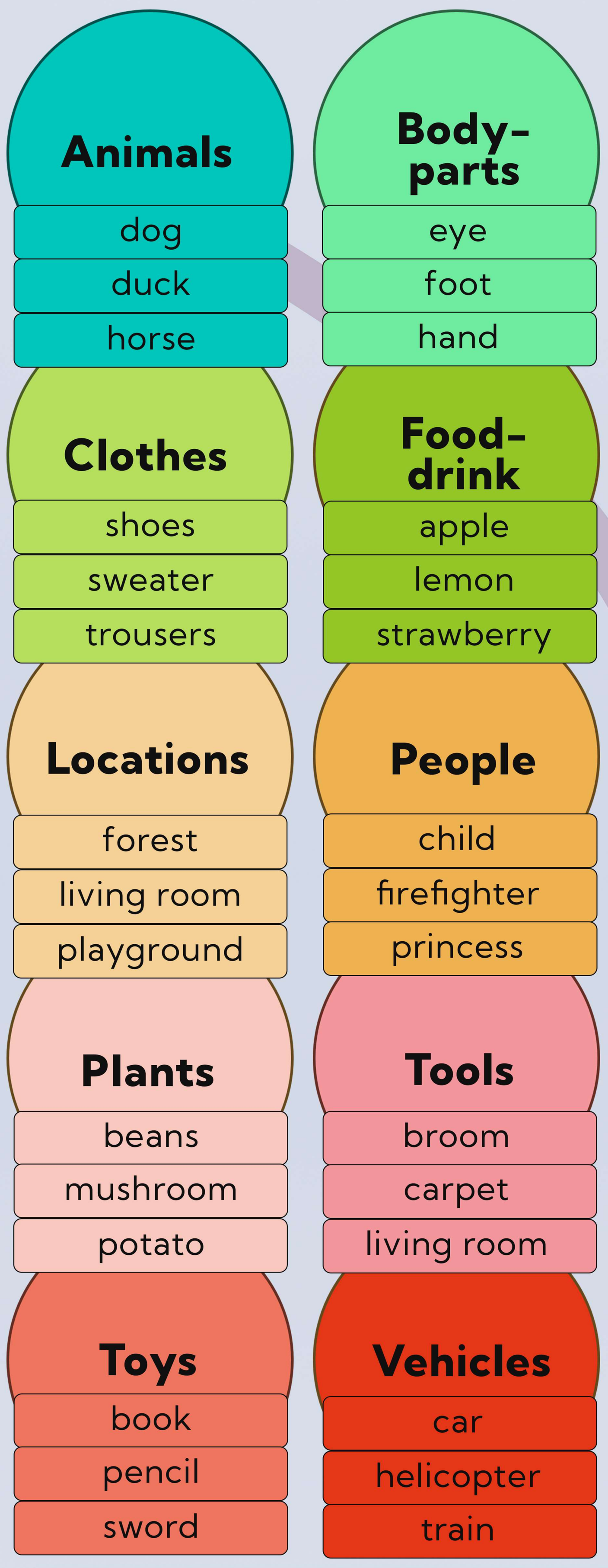
Developing a concept feature norm for studying concept representations in children aged 6–9-years-old

PS3-12

Alex Ilyés[1,2,3], Borbála Brosig[4,2], Anna Noémi Illés[4,2], Donát Keszthelyi[4,2], Éva Kovaliczky[4], Boróka Ferencz[4,2], Attila Keresztes[2,3]

ilyes.alex@ttk.hu

1. Doctoral School of Psychology, ELTE Eötvös Loránd University, Hungary; 2. Brain Imaging Centre, HUN-REN Research Centre for Natural Sciences, Hungary; 3. Institute of Psychology, ELTE, Hungary; 4. Faculty of Education and Psychology, ELTE, Hungary



Introduction

We aim to collect at least **30 responders** for a single concept (**total Nconcept = 300**) each producing a minimum of **five different** features [1,2].

We created a **child-friendly website** to collect data online [3] from all over Hungary.

We selected concepts for a **'mini-norm' (selected Nconcept = 30)** to pilot the processing pipelines. Each depicted analysis is data-driven and exploratory.

Participants

We recruited 48 Hungarian children (23 females, Mage = 7.63, SD = 0.94) from a single elementary school.

Methods

On average children produced 14.42 features (SD = 5.48; range = 6.7 – 35.7) with a minimum of five feature per concept

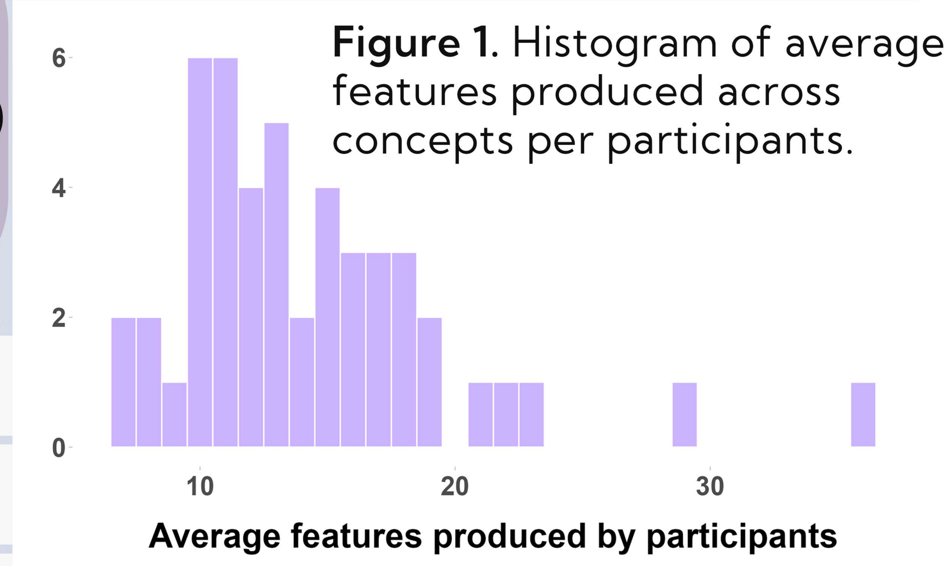


Figure 1. Histogram of average features produced across concepts per participants.

Winged creature

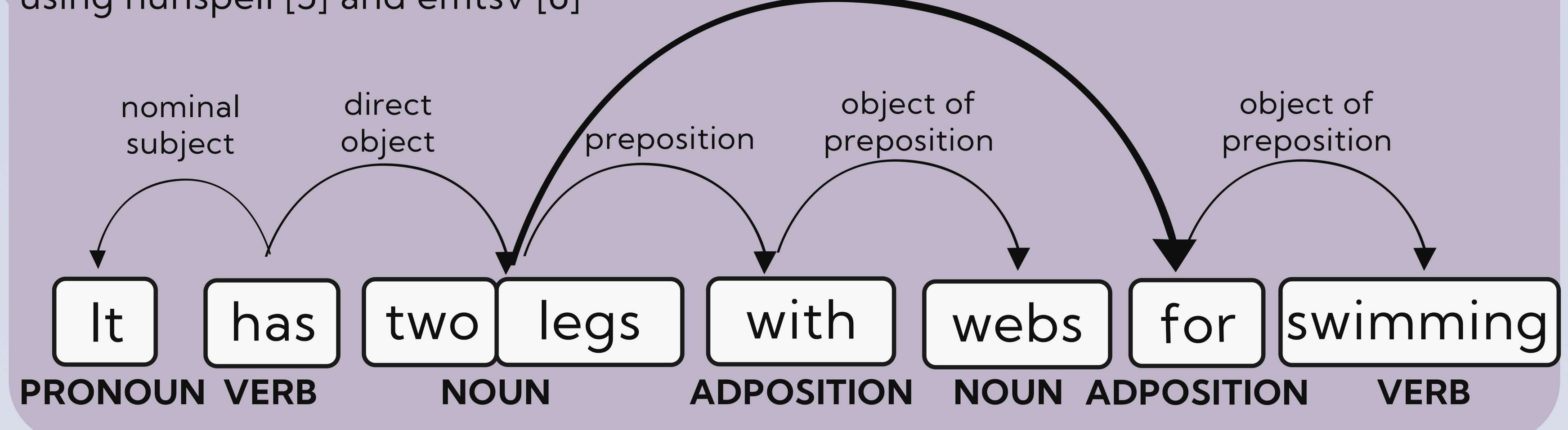
It can fly

It usually lives in lakes

It has two legs with webs for swimming

It has feathers

We extracted features by grammatical co-occurances using hunspell [5] and emtsv [6]



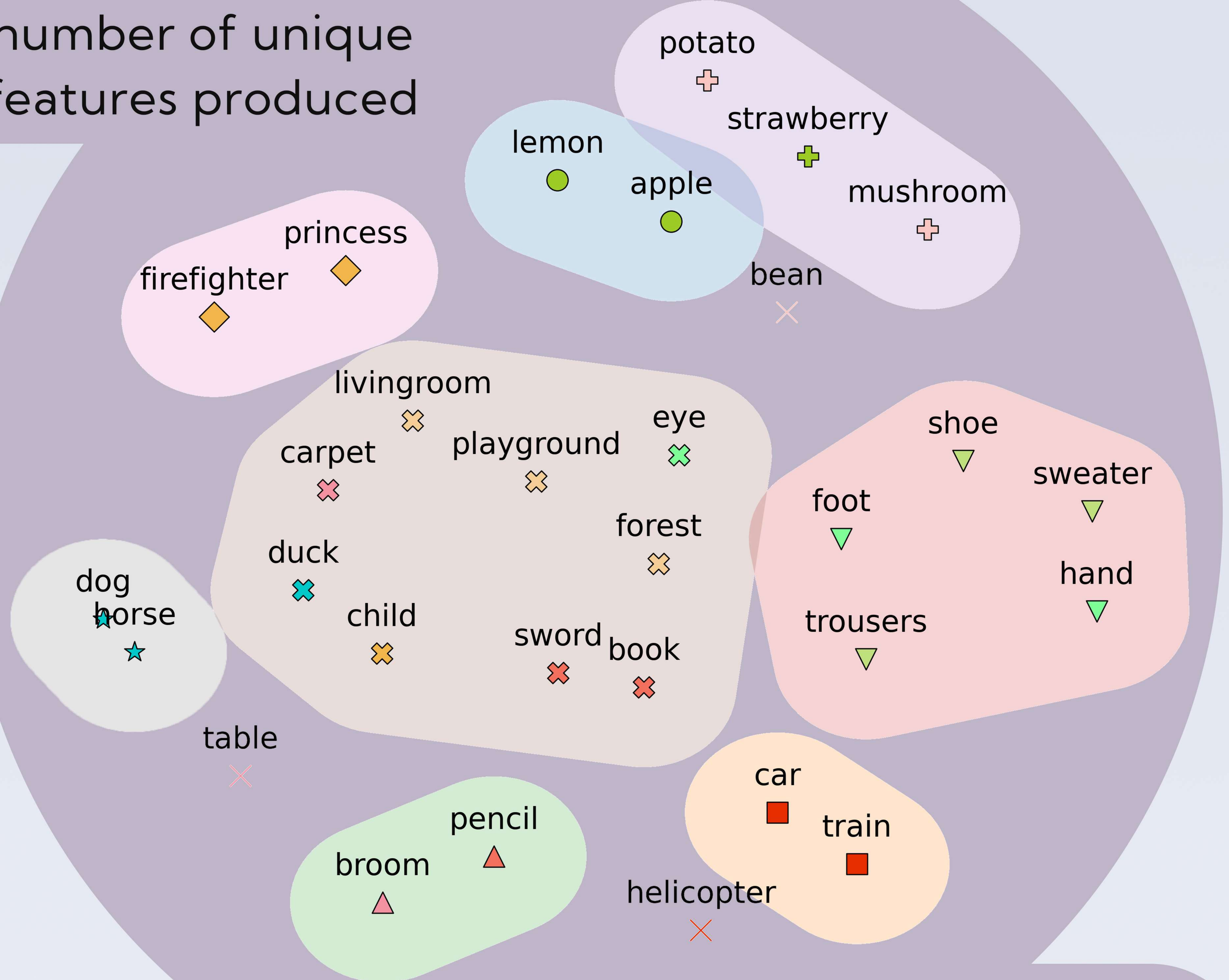
verb + direct object: HAS_LEG

quantity + object: TWO_LEG

object + prep. object: HAS_WEB_LEG

action verb: CAN_SWIM

We created an N-dimensional vector space, where N is the number of unique features produced



We plotted two UMAP-based [7] principal components and clusters via DBSCAN [8]

Results

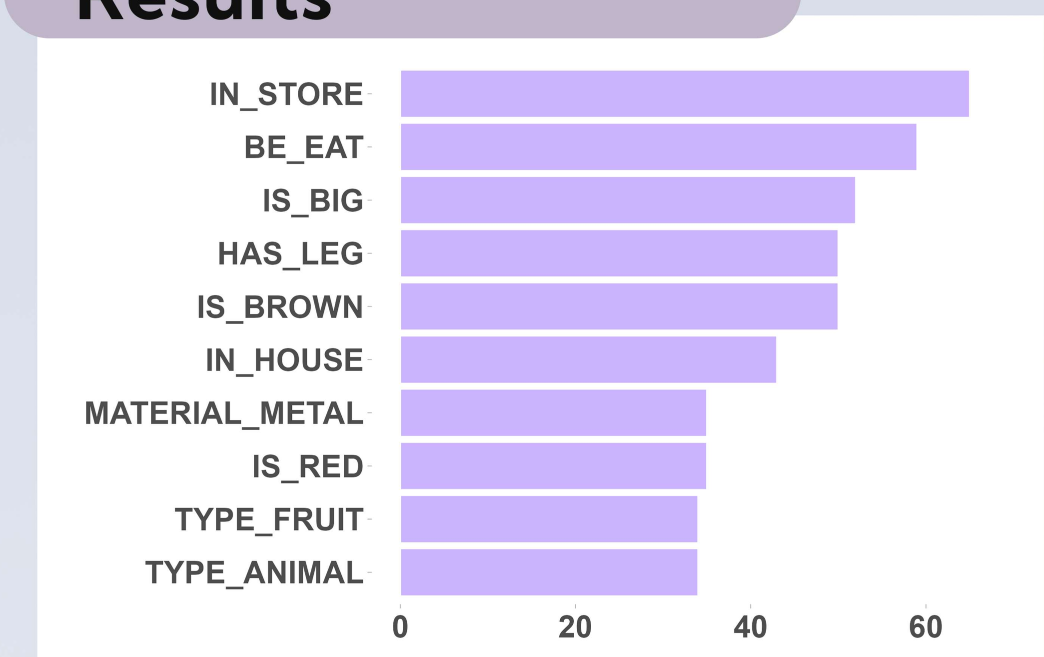


Figure 2. Bars represent the frequency (x-axis) of the ten most-produced features in the dataset.

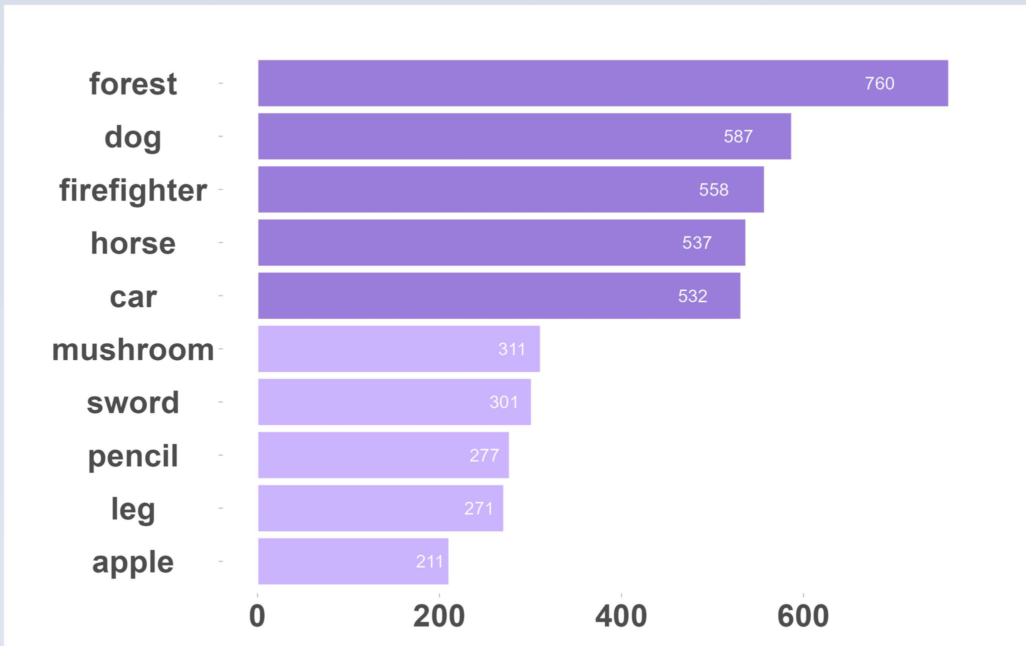
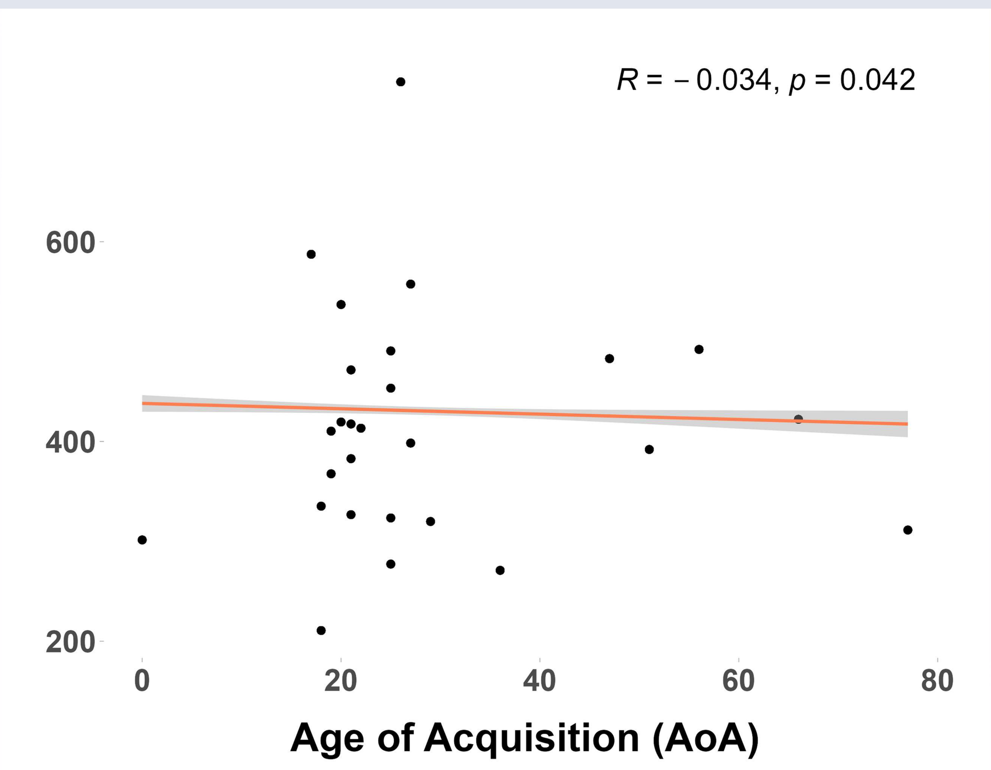


Figure 3. Bars represent the five most and five least "semantically rich", i.e., the amount of unique features collected and predicted by the following equation (x-axis):

$$S^{*}_{obs} + A^{**}\frac{Q_1^{2***}}{2Q_2}$$

* number of features in total
** sample size–1/sample size
*** number of features mentioned by indexed number of subjects (based on [9])



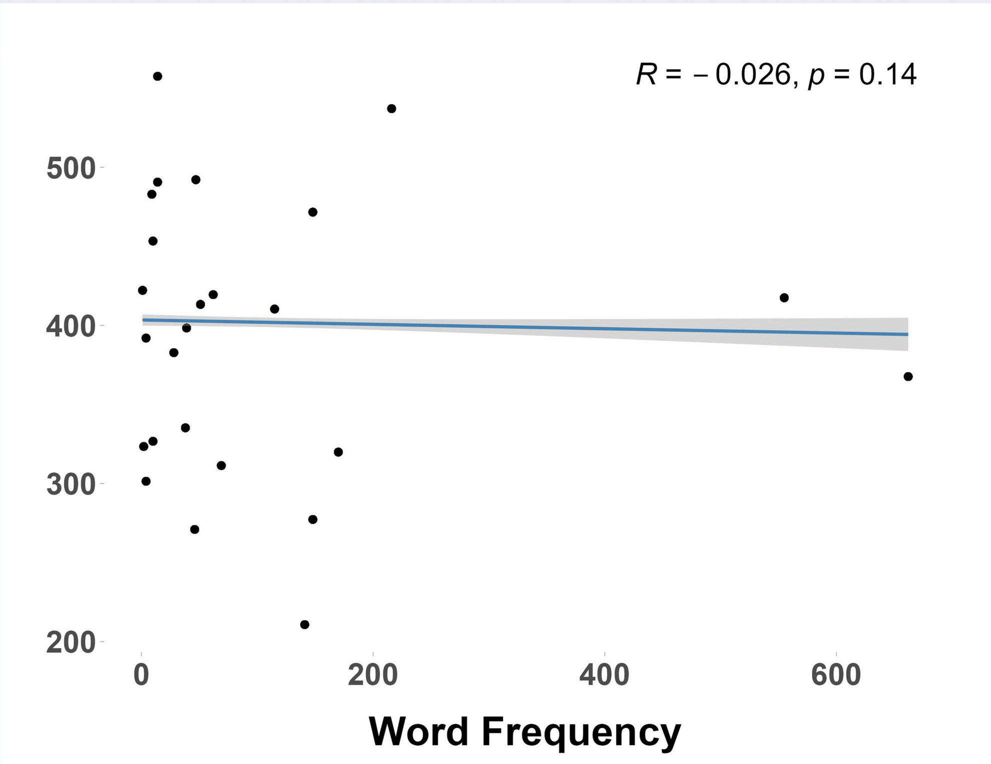


Figure 4-5. Pearson correlations between semantic richness and age-of-acquisitions [10] (above) and frequency [11] (below) of the concepts.

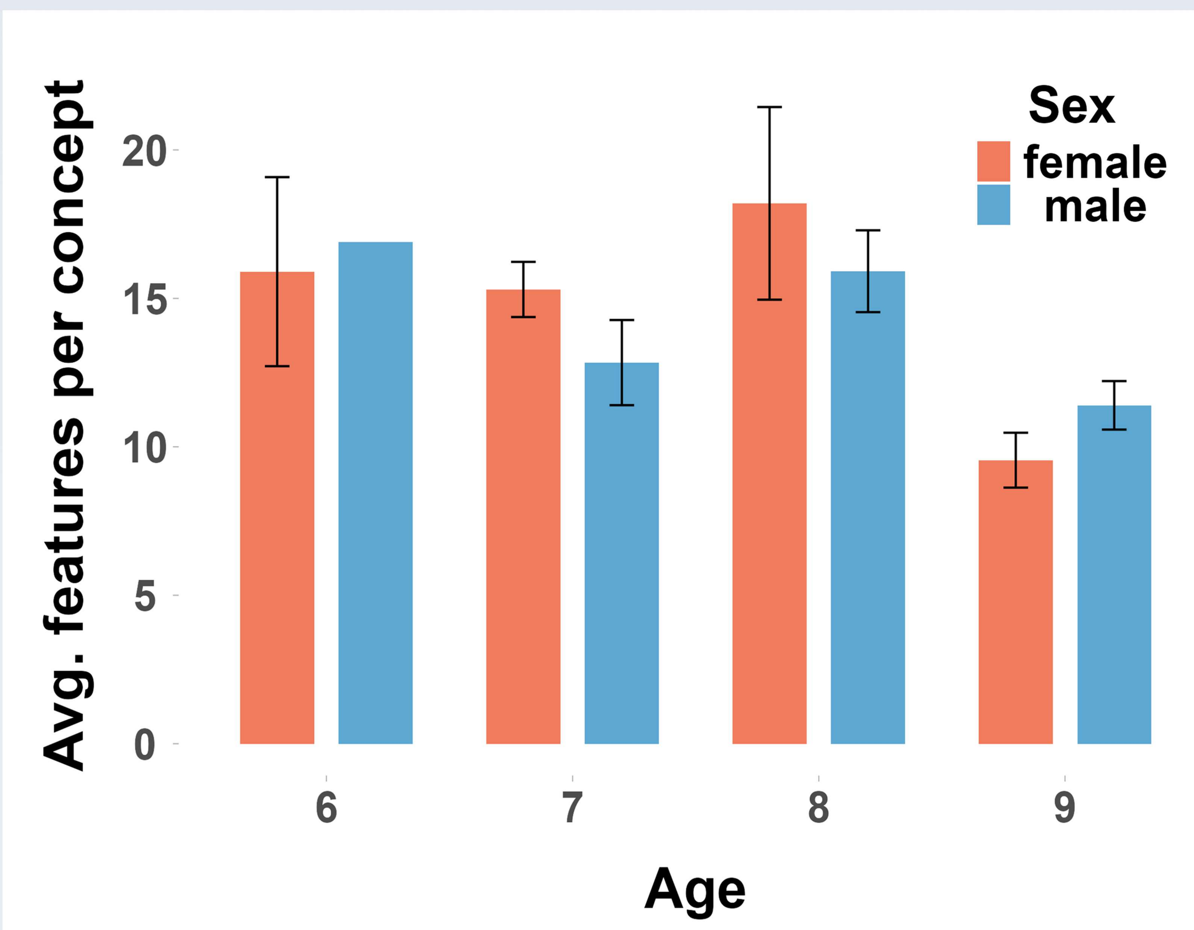


Figure 6. Age in years is plotted on the x-axis, while participant-specific feature count averaged across seen concepts is on the y-axis, males in blue and females in red. Error bars depict standard error.

We found no effects of age (F = 1.49, p = 0.23) or sex (F = 1.15, p = 0.29) in a two-way ANOVA predicting feature production rate.

Take-away

Acquiring concept features from children aged 6-9 years is feasible on a large scale.

Almost unequivocally enjoyed teaching Mimó.

Children produced surprisingly high amount of features during production which resulted in a well-clustered network, although with many overlaps between canonical categories.

Neither age, sex, or concept frequency had an effect on feature count, early acquired concepts had marginally higher feature responses.

References

This work was supported by the Hungarian National Research, Development and Innovation Office – NKFIH [FK 146496], a Bolyai János Research Scholarship of the Hungarian Academy of Sciences, and a Max Planck Partner Group by the Max Planck Society awarded to Attila Keresztes.

1. Deng, Y. et al. A Chinese Conceptual Semantic Feature Dataset (CCFD). Behav. Res. Methods 53, 1697–1709 (2021).
2. Devereux, B. J., Tyler, L. K., Geertzen, J. & Randall, B. The Centre for Speech, Language and the Brain (CSLB) concept property norms. Behav. Res. Methods 46, 1119–1127 (2014).
3. Gholizadeh, N., Saadatfar, H. & Hanafi, N. K-DBSCAN: An improved DBSCAN algorithm for big data. J. Supercomput. 77, 6214–6235 (2020).
4. Heredia, C., Moreno, S., Canessa, E. & Chaigneau, S. WordListsAnalytics: An R package for multiple data analysis of Property Listing Tasks. SoftwareX 27, 101842 (2024).
5. Hovhannisyan, M. et al. The visual and semantic features that predict object memory: Concept property norms for 1,000 object images. Mem. Cognt. 49, 712–731 (2021).
6. Indig, B. et al. One format to rule them all – The emtsv pipeline for Hungarian. in Proceedings of the 13th Linguistic Annotation Workshop (eds. Friedrich, A., Zeyrek, D. & Hoek, J.) 155–165 (Association for Computational Linguistics, Florence, Italy, 2019). doi:10.18653/v1/W19-4018.
7. Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. Age-of-acquisition ratings for 30,000 English words. Behav. Res. Methods 44, 978–990 (2012).
8. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426 (2020).
9. McRae, K., Cree, G. S., Seidenberg, M. S. & Mcnorgan, C. Semantic feature production norms for a large set of living and nonliving things. Behav. Res. Methods 37, 547–559 (2005).
10. Oravecz C. A Magyar nemzeti szövegtár. [The Hungarian National Corpus. From corpus linguistics to neural networks]. vol. A korpusznyelvészettől neurális hálókig (Nyelvtudományi Kutatóközpont, Budapest, 2021).
11. Trón, V. et al. Hunmorph: open source word analysis. in Proceedings of the Workshop on Software 77–85 (Association for Computational Linguistics, USA, 2005).